ARMY RESEARCH LABORATORY

An Application of Statistics to Algorithm Development in Image Analysis

Barry Bodt, Philip David, and David Hillis

ARL-TR-2333 August 2001

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5425

ARL-TR-2333 August 2001

An Application of Statistics to Algorithm Development in Image Analysis

Barry Bodt, Philip David, and David Hillis

Computational and Information Sciences Directorate

Sponsored by

Ballistic Missile Defense Organization

7100 Defense, Pentagon

Washington, DC 20301-7100

Abstract

We gathered and processed infrared (IR) images taken by a quantum well infrared photodetector (QWIP) camera to support the development of an algorithm for airborne target detection. This report addresses the statistical aspects of this effort. Digitized images consisted of background clutter alone and background clutter with a small aircraft target. We used a template-matching technique to detect the small aircraft amidst the background clutter. We propose a reasonable statistic for use in our algorithm. Our primary focus is to describe the experimental design and analysis used in determining parameter settings for the detection algorithm and to report those results.

Contents

Figures

Tables

1. Introduction

We gathered and processed infrared (IR) images taken by a quantum well infrared photodetector (QWIP) camera to support the development of an algorithm for airborne target detection. This report addresses the statistical aspects of this effort. Digitized images consisted of background clutter alone and background clutter with a small aircraft target. We used a template-matching technique to detect the small aircraft amidst the background clutter. We propose a reasonable statistic for use in our algorithm. Our primary focus is to describe the experimental design and analysis used in determining parameter settings for the detection algorithm and to report those results.

Multiple tapes of IR imagery were collected. Two aircraft flew over the same area at different altitudes. A Cessna aircraft was used as a surrogate target vehicle. The Cessna's path crossed that of the observing aircraft, flying at a greater altitude with its IR sensor pointing straight down. Each time the two aircraft crossed, the Cessna's flight was recorded by the continuous QWIP imagery. Figures 1 through 5 show representative imagery. One hundred fifty-two images were selected from these tapes, showing the Cessna at different apparent ranges and aspect angles and against a variety of natural background clutter. Fifty additional images without the Cessna were also selected. These additional images included more confusing man-made clutter. All images were digitized with 512×512 8-bit pixels with grayscale values ranging from white (255) to black (0).

Our goal was to develop a reliable algorithm for automatic detection of the Cessna.

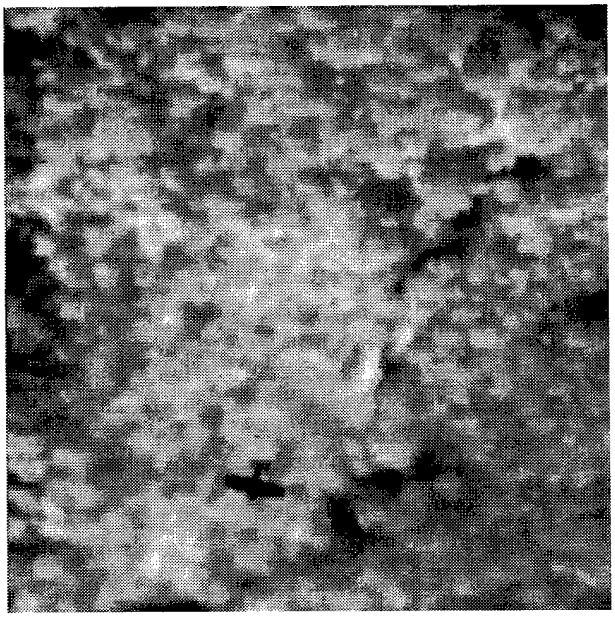


Figure 1. QWIP IR. Image of Cessna against natural background. Provided by Naval Research Laboratory.

Figure 2. QWIP IR. Image of Cessna against natural background. Provided by Naval Research Laboratory.

Figure 3. QWIP IR. Image of Cessna against natural background. Provided by Naval Research Laboratory.

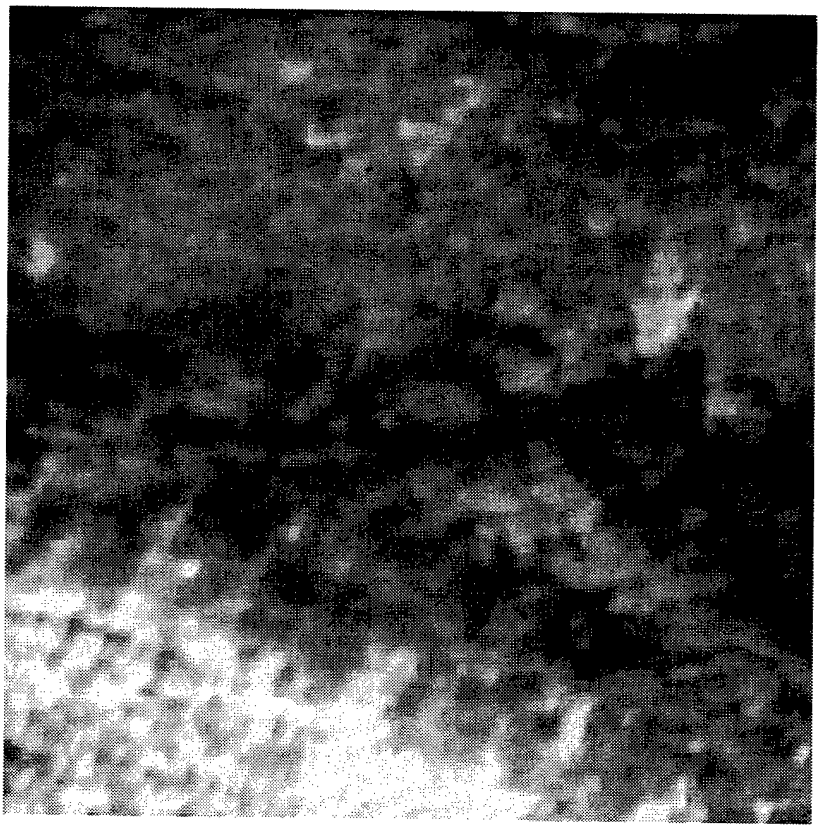


Figure 4. QWIP IR. Image of man-made background. Provided by Naval Research Laboratory.

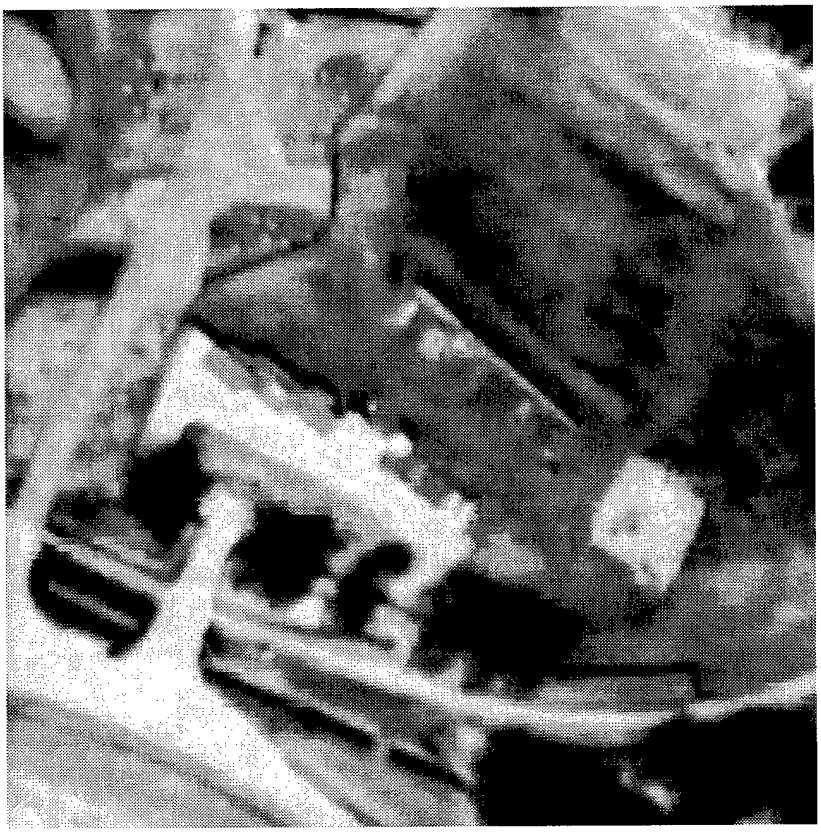
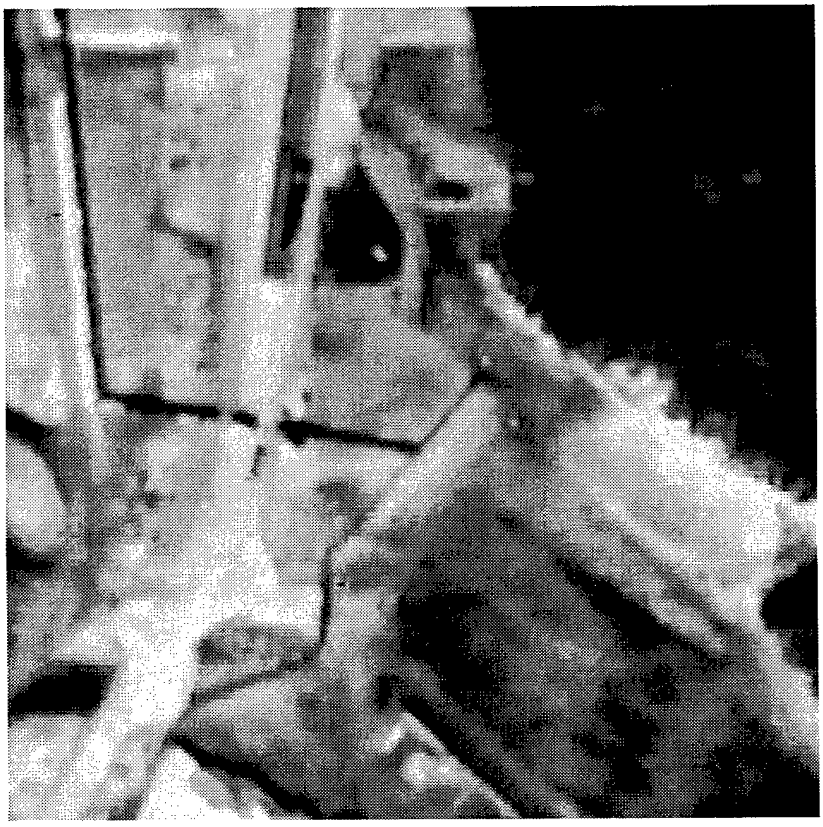


Figure 5. QWIP IR. Image of man-made background. Provided by Naval Research Laboratory.



2. Template Matching

Before introducing the complete algorithm, we briefly discuss template matching, a key component of that algorithm. As the name suggests, template matching involves overlaying a template on the image and testing to determine if the image and the template match. A match is said to occur if there is sufficient agreement between the edges of a target template and the edges of the image. Since the locations of the template edges are known, the interior and exterior of the template are completely defined. In the images examined, the Cessna was invariably quite cold. If the template and target were exactly aligned, we would expect pixel-grayscale values on the interior of the template to be small (darker) and relatively homogeneous. Pixel-grayscale values exterior to the template (in the background) would tend to be larger (lighter) and less homogeneous. Consider an edge of the aligned template and target. The difference across this edge between an exterior pixel value and a neighboring interior pixel value would tend to greatly exceed zero. This property of an edge serves as the basis for our decision regarding whether the image and template match.

We wish to parlay this property into a reasonable algorithm to detect the presence of the Cessna in our image. A formal hypothesis test is not developed. Rather, we offer some justification for one of the test statistics tried and then concentrate on determining optimum parameter values for our algorithm based on a learning sample of images.

The statistic arises as follows. Thirty-eight ordered pairs are formed in fixed relation to the template, with each pair consisting of an exterior point and a neighboring interior point. (Pairs are numbered consecutively around the template.) Let E_j and I_j denote the grayscale values of the j^{th} exterior and interior points, respectively. Ignoring any dependence, consider the pixels in two adjacent ordered pairs as samples of size 2 from the interior and exterior regions. Equation (1) can be thought of loosely in terms of rationale for analysis of variance in that we are attempting to compare variation across groups with variation within groups. We take several liberties. The last term represents within variability, based only on the pixels internal to the template. If we substitute an absolute value operator for the traditional square, then $|I_j - I_{j+1}|$ is the sum of absolute deviations from the sample mean $(I_j + I_{j+1})/2$. Doubling this value accounts for the contribution to within variation made by the external pixels under the assumption that the variation is similar. For between variation, consider that $(E_j + E_{j+1})/2$ is an estimate for the external mean and that $(E_j + E_{j+1} + I_j + I_{j+1})/4$ is an estimate for the grand mean. The sum of squares between is analogous to four times the absolute difference between these two terms, leaving $|(E_j + E_{j+1}) - (I_j + I_{j+1})|$. Finally, because of the interest in a one-sided alternative with external values being expected to exceed internal values, the absolute value is dropped. We are left with a heuristically justified difference in eq (1), which will assume large

values when the two pairs each straddle a template edge that is aligned with the target.

$$(E_j - I_j) + (E_{j+1} - I_{j+1}) - 2\,|I_j - I_{j+1}|\ . \tag{1}$$

Two related statistics that we considered are given as equations (2) and (3). Equation (2) is reasoned along the lines of (1), but *within variation* is computed separately interior and exterior to the template. Equation (3) makes only one comparison across a potential edge and uses the neighboring j+1st internal pixel for a within-variation estimate. The multiplication by two adjusts the scale of equation (3) to that of equations (1) and (2).

$$(E_j - I_j) + (E_{j+1} - I_{j+1}) - (|E_j - E_{j+1}| + |I_j - I_{j+1}|) \tag{2}$$

$$2(E_j - I_j - |I_j - I_{j+1}|) \tag{3}$$

3. Proposed Algorithm

The previous discussion suggests why proposed statistics might reveal an edge. To complete the test of template and target alignment, however, we must look at the entire template boundary as a function of scale and orientation. For fixed scale and orientation, the template boundary evaluation is supported by $\{[(E_j, I_j), (E_{j+1}, I_{j+1})]: j = 1, 38\}$, with the first point reused for $j = 39$. Thirty-eight individual tests based on (1) to (3) record a response if the statistic exceeds some threshold level τ. For the edges of the template to be said to match the target, the collective responses for edges around the target must exceed a second threshold, T. These two parameters, τ and T, determine whether a proposed template aligns with the target, that is, whether the target is detected.

Equally important is determining what templates should be proposed. An image of 512×512 pixels is searched for a subset that collectively yields edges consistent with the target. The subset is a collection of 38 pixel pairs with fixed location relative to a prospective center point for the target. For the template to match the target, both must have approximately the same angular orientation in the view considered. Thus, many orientations must be tried to achieve coverage of possible orientations. Additionally, the template must be adjusted for various ranges. A target overlain by a template with the true center point and the correct orientation but very different scale will likely not be detected. Thus, a sufficiently fine resolution of scales must also be taken into account. Considering, for example, each pixel as a potential center point, 36 orientations and 8 scales yields a set of 75,497,472 possible templates to be proposed.

To reduce the number of computations, we perform the following preprocessing. At what we will refer to as level 1, we consider points as potential template centers only if their grayscale value ≤ 5. This is reasonable since the target center shows up nearly black on the images we have. For each of these center points, we take a sparse sample of orientation $\times$ range. Azimuths are taken every 30° instead of every 10°. Two rather than eight scales are tested. Only seven ordered pairs about the template are considered for an edge test. For each center point passing this reduced edge test, we also flag the eight adjacent neighbors. (Some multiple counting would be expected.) Thus, the preprocessing would result in a set of potential center points, along with their nearest neighbors, that had passed the grayscale screen and reduced edge test screen for some orientation and scale.

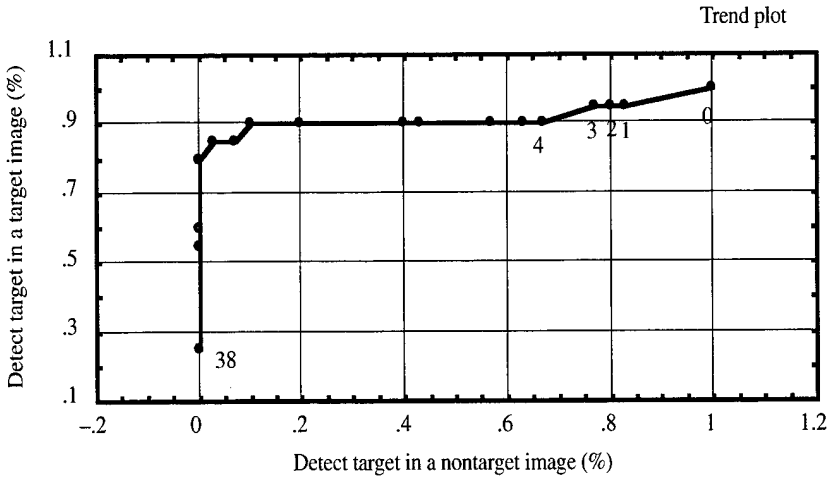
The proposed algorithm detects a target if a pixel passing through preprocessing for a specific orientation and scale results in more than T edge tests about the template exceeding the threshold τ.

4. Strategy for Optimization

A response surface approach was used to fix the four parameters of the model (τ, T, number of scales, number of orientations) and to determine which of the three test statistics performed best. Two responses are considered: the maximum probability of detection while maintaining zero false alarms (Max_P(D)) and the area under a receiver operator curve (ROC), ROC_Area. For each setting of the algorithm parameters, we evaluated two sets of images—those containing a target and those not containing a target. Consider an orientation, scale, and edge test threshold τ. The ROC arises by recording the proportion of images whose best pixel passes at least 0, 1, 2, ... , 38 edge tests. All images, with or without targets, will pass zero edge tests and earn a proportion 1. See figure 6. Few images with targets and, ideally, no images without targets would pass all 38 edge tests to earn proportions of approximately zero. The goal in examining the ROC is to determine the number of edge tests (0 to 38) that make the algorithm sufficiently sensitive to detect a target without many false alarms. An area under the curve of near 1 is indicative of such a test.

In our computer experimentation, we examine the impact of orientation, scale, and τ for each of three proposed statistics using the ROC_Area and Max_P(D), with an eye toward optimizing those responses. The threshold T for the number of edge tests required to conclude detection is determined from examining near-optimum ROCs.

Figure 6. A promising ROC.



5. Learning Sample Experiments

Twenty images containing targets were selected. Portions of these images with no target present were used as nontarget images (whole image less a 60×60-pixel area about the target). Ten additional pure clutter images were also used. Thus, the target images were considered to number 20 and the nontarget images, 30 for the training testing.

5.1 Learning Set Analysis Phase I

The computer experiments for the initial test region were run with the factors and levels given in table 1. The threshold τ is the amount a chosen edge test statistic must exceed for one to conclude that an edge is present. The orientation is the number of steps taken over 360° for template positioning. For example, 24 steps correspond to 15°. Stepping is performed relative to a random start orientation. Scale indicates the number of uniformly spaced divisions over the interval of target ranges considered. The statistic refers to the expressions introduced in equations 1 to 3. We ran a complete factorial, using four- and five-way interactions as error.

Analysis of variance indicated that only the first three factors of table 1 were significant. See the main effects plot in figure 7. Interestingly, the statistic used was not significant. No two- or three-way interactions were found to be significant. See figure 8. This suggests that the algorithm parameters could be adjusted independently. A regression model of ROC_Area as a function of only the three significant main effects explained 79.2 percent of the variation and satisfied the usual regression analysis assumptions. Figures 9 and 10 show the best and worst ROC over the factor levels tested. Figure 9 corresponds to resolution on orientation of 24, resolution for scale of 6, and τ = 75, whereas figure 10 is based on values of 8, 2, and 135 for orientation, scale, and τ, respectively.

Similar results were seen with the use of MaxP(D) as the response. The same three factors are significant and a well-behaved regression model explains 75.5 percent of the variation.

Both responses showed room for improvement. The analysis suggested that each might be moved closer to optimum by (1) increasing the resolution of scale, (2) increasing the resolution of orientation, and (3) lowering the threshold for detection of an edge.

Table 1. Initial test region factor levels.

Factor	Levels
τ	75, 90, 105, 120, 135
Orientation	24, 12, 8
Scale	2, 4, 6
Statistic	1, 2, 3

Figure 7. Main effects plot with response ROC_Area.

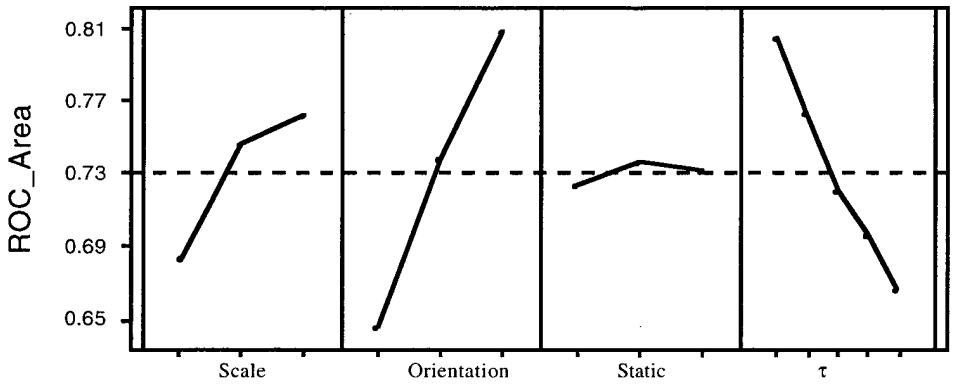


Figure 8. Interaction plot with response ROC_Area

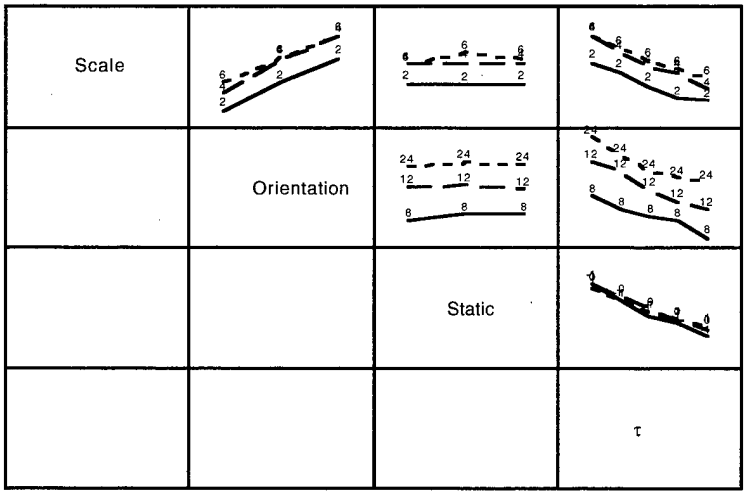


Figure 9. Best ROC.

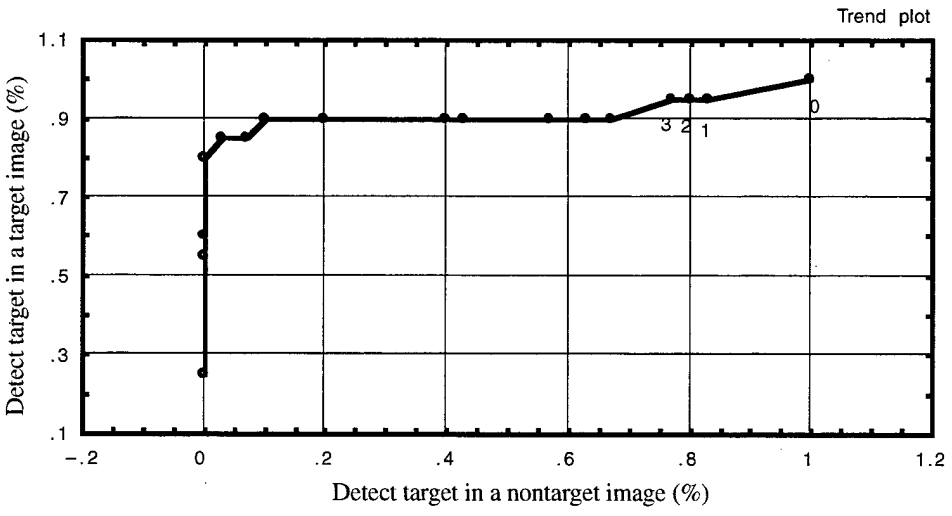
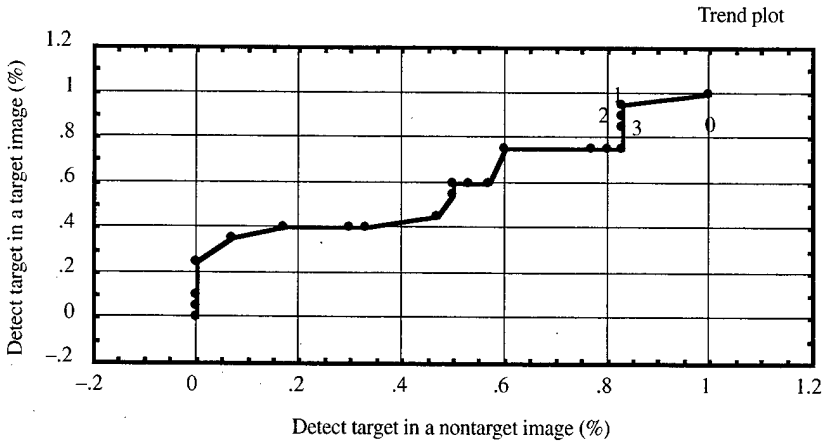


Figure 10. Worst ROC.



A second group of experiments was run based on what we learned in the first phase. Table 2 lists the factors and levels used in this second set of tests. The three statistics (1) to (3) were all retained for this test.

The analysis proceeded similarly to that of phase I. Again, only the three main effects drove the model. No significant interactions were present. Figure 11 shows the main effects. From this graph we can see that the ROC_Area range in which the testing is done is much higher than in the first phase of testing, all means showing areas in excess of 0.91. We also see that the influence of τ seems to level off as the ROC_Area nears 0.95. Values of τ less than or equal to 60 yield comparable results. A best-fitting regression model does include a square term for τ, so the apparent curvature is probably real. The explained variation over this more narrow range of responses is 85.8 percent.

Boxplots incorporating the data from phases I and II were constructed to show the distribution of response measures as a function of the three significant factors. See figures 12 to 14. We wished to determine the minimum scale and orientation and maximum τ that could be used to provide a near optimal ROC. Algorithm speed was a concern. More scales and orientations improve detection at the cost of computational efficiency. A larger differential across an edge was required for τ so that we could be sure that the apparent edge was not just a difference in clutter values. In examining these three figures, we conclude that there should be 8 resolutions of scale, 36 orientations, and a value of about 60 for τ. A refinement for orientation was made after looking at specific plots of individual values. In our judgment, 24 orientations would suffice.

The parameter for our algorithm not yet established is the threshold for the number of edge tests out of the 38 that must be passed for a detection to be said to occur. Our approach to finding this threshold was to examine the ROCs for which the ROC_Area was very high and consistent with our optimal settings for the other parameters. For each of these we recorded the number of edge tests passed corresponding to the Max_P(D). The results did not vary greatly, all being 34, 35, or 36. We chose to set the threshold for the number of edge tests passed at 35.

Table 2. Phase II Test region factor levels.

Factor	Levels
τ	40, 50, 60, 70, 80
Orientation	72, 36, 24
Scale	6, 8, 10
Statistic	1, 2, 3

Figure 11. Main effects plot.

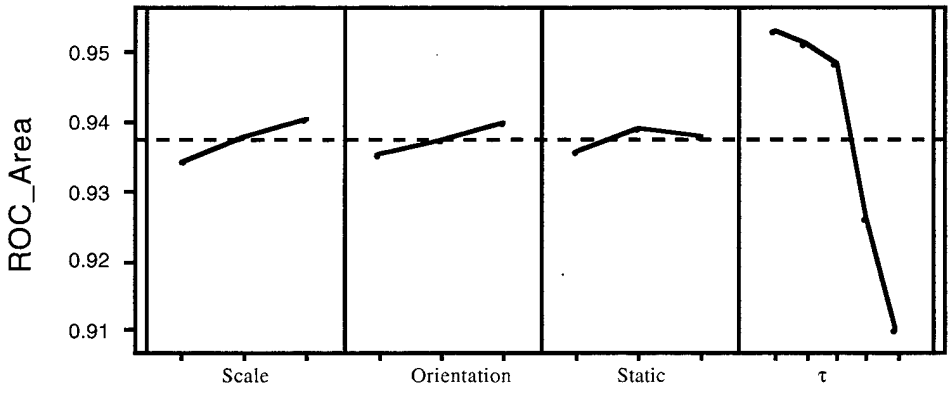


Figure 12. Boxplots of ROC_Area by scale.

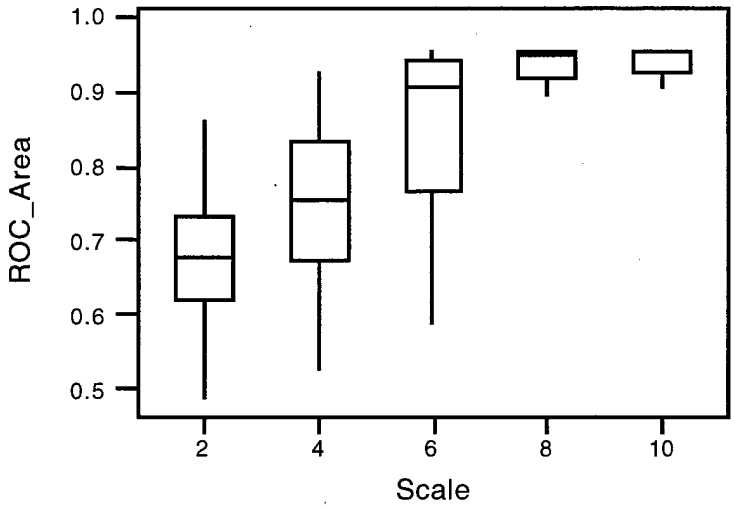


Figure 13. Boxplots of ROC_Area by orientation.

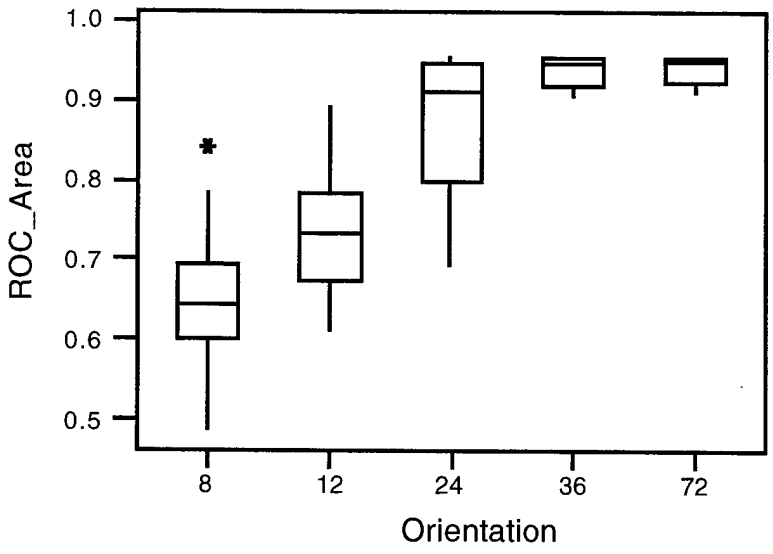
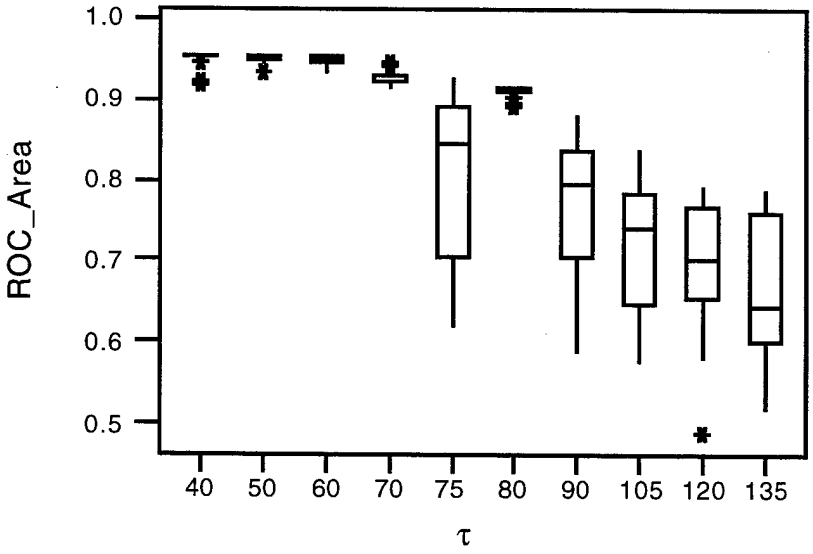


Figure 14. Boxplots of ROC_Area by τ.



6. Test Sample Experiment

A final test was run to confirm the choice of parameters for our model: τ = 60, orientation = 24, scale = 8, and edge test threshold = 35. We chose to test one level to either side of the optimum. The threshold values of 34, 35, and 36 were considered for each of the 27 experimental combinations from a factorial crossing of the factors in table 3.

One hundred thirty-two images containing targets were selected for the test phase. Portions of these images with no target present were used as nontarget images (whole image less a 60 × 60-pixel area about the target). An additional 40 pure clutter images were also used. Thus the target images were considered to number 132 and the nontarget images 172 for the test phase. These images were similar but not identical to those used in the previous phases.

Figure 15 shows the results of the 27 trials formed from the factor crossings of table 3. Results are given in terms of the two responses, ROC_Area and Max_P(D). As expected, the chosen parameter settings produced results in the midrange for each response over the 27 trials for the larger test data set. All of the runs over the test data set are included in table 4. It is clear from scanning this table that all experimentation is in a region showing good results for both response measures. The ideal tradeoff between probability of detection and computational efficiency will be dependent upon the application in which it is used.

Table 3. Test phase factor levels.

Factor	Levels
τ	50, 60, 70
Orientation	36, 24, 18
Scale	6, 8, 10

Figure 15. Algorithm performance.

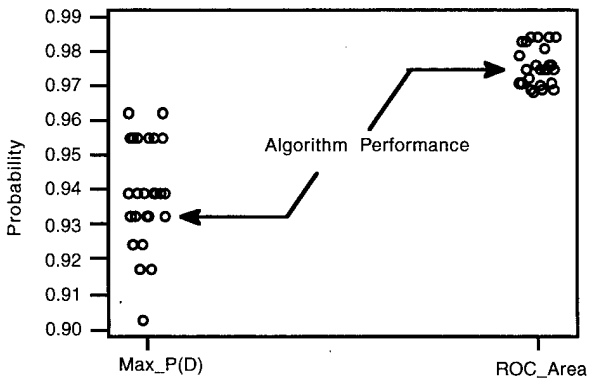


Table 4. Summary algorithm performance.

Max_P(D)	ROC_Area	Scale	Orientation	τ	Threshold
0.962	0.984	6	36	50	36
0.955	0.976	6	36	60	35
0.939	0.971	6	36	70	34
0.955	0.984	6	24	50	36
0.932	0.976	6	24	60	35
0.917	0.969	6	24	70	35
0.939	0.981	6	18	50	36
0.939	0.976	6	18	60	35
0.917	0.968	6	18	70	34
0.939	0.984	8	36	50	37
0.939	0.975	8	36	60	35
0.932	0.971	8	36	70	35
0.955	0.983	8	24	50	36
0.932	0.975	8	24	60	35
0.932	0.972	8	24	70	34
0.955	0.984	8	18	50	36
0.939	0.975	8	18	60	35
0.902	0.969	8	18	70	34
0.962	0.984	10	36	50	36
0.939	0.976	10	36	60	35
0.924	0.970	10	36	70	35
0.955	0.983	10	24	50	36
0.939	0.975	10	24	60	35
0.917	0.969	10	24	70	35
0.955	0.979	10	18	50	36
0.932	0.975	10	18	60	35
0.924	0.971	10	18	70	34

7. Discussion

In this paper, we used a mix of engineering practice and statistics to try to arrive at an algorithm for airborne target detection. Admittedly, the approach was somewhat informal from a statistical perspective, but it seemingly produced good results. Extension to other clutter backgrounds would require additional testing.

We did consider more traditional approaches, specifically a paired-t signed rank test and a Wilcoxon matched-pairs signed rank test. In each instance, we must overlook the spatial correlation present among neighboring interior and neighboring exterior points. Threshold T is analogous to the significance level of the test. Threshold τ is analogous to a general mean difference to be exceeded. Setting these values with respect to our optimization goals requires reasoning along the lines used in this paper. With assumptions for these procedures in question and with no real advantage to the general modeling approach, we settled on the approach presented here.

8. Acknowledgment

The authors would like to thank the Ballistic Missile Defense Organization and the Naval Research Laboratory for providing the data and this opportunity to explore target-detection algorithms.

9. References

Baras, J.S., and D.C. MacEnany (1992), "Model based ATR algorithms based on reduced target models, learning and probing," Proc. 2nd of the Second Automatic Target Recognizer Systems and Technology Conference, 1992.

Bienenstock, D., D. Geman, S. Geman, and D.E. McClure (1990), *Phase II Technical Report, Development of Laser Radar ATR Algorithms*, Contract No. DAAL02-89-C-0081, Night Vision and Electronic Sensors Directorate (CECOM), Ft. Belvoir, VA.

Nguyen, D.M. (1990), An Iterative Technique for Target Detection and Segmentation in IR Imaging Systems, Ft. Belvoir, VA.

Severson, W.E. (1996), The SSV Vehicle's FLIR Target Detection Capability, Lockheed Martin Astronautics, Contract No. DASG60-95-C-0062, U.S. Army Space & Strategic Defense Command.

Distribution

Admnstr
Defns Techl Info Ctr
ATTN DTIC-OCP
8725 John J Kingman Rd Ste 0944
FT Belvoir VA 22060-6218

DARPA
ATTN S Welby
3701 N Fairfax Dr
Arlington VA 22203-1714

Ofc of the Secy of Defns
ATTN ODDRE (R&AT)
The Pentagon
Washington DC 20301-3080

AMCOM MRDEC
ATTN AMSMI-RD W C McCorkle
Redstone Arsenal AL 35898-5240

US Army TRADOC
Battle Lab Integration & Techl Dirctrt
ATTN ATCD-B
FT Monroe VA 23651-5850

BMDO
ATTN L Lome
7100 Defense Pentagon
Washington DC 20301-7100

Hdqtrs Dept of the Army
ATTN DAMO-FDT
400 Army Pentagon Rm 3C514
Washington DC 20301-0460

US Military Acdmy
Mathematical Sci Ctr of Excellence
ATTN MADN-MATH MAJ M Johnson
Thayer Hall
West Point NY 10996-1786

Dir for MANPRINT
Ofc of the Deputy Chief of Staff for Prsnnl
ATTN J Hiller
The Pentagon Rm 2C733
Washington DC 20301-0300

SMC/CZA
2435 Vela Way Ste 1613
El Segundo CA 90245-5500

US Army ARDEC
ATTN AMSTA-AR-TD
Bldg 1
Picatinny Arsenal NJ 07806-5000

US Army Info Sys Engrg Cmnd
ATTN AMSEL-IE-TD F Jenia
FT Huachuca AZ 85613-5300

US Army Natick RDEC Acting Techl Dir
ATTN SBCN-T P Brandler
Natick MA 01760-5002
US Army Simulation Train & Instrmntn
Cmnd
ATTN AMSTI-CG M Macedonia
ATTN J Stahl
12350 Research Parkway
Orlando FL 32826-3726
US Army Tank-Automtv Cmnd RDEC
ATTN AMSTA-TR J Chapin
Warren MI 48397-5000

Nav Rsrch Lab
ATTN Code 7604 J Goldspeil
4555 Overlook Ave SW Bldg 209 Rm 124
Washington DC 20375-5352

Nav Rsrch lab Sensor & Data Analysis Sect
ATTN Code 5621 F Bucholtz
Washington DC 20375-5000

Nav Surfc Warfare Ctr
ATTN Code B07 J Pennella
17320 Dahlgren Rd Bldg 1470 Rm 1101
Dahlgren VA 22448-5100

Univ of Texas at Austin
Inst for Advncd Tchnlgy
PO Box 202797
Austin TX 78720-2797

Hicks & Assoc Inc
ATTN G Singley III
1710 Goodrich Dr Ste 1300
McLean VA 22102

Distribution

US Army Rsrch Lab
ATTN AMSRL-CI-CT A Neiderer
ATTN AMSRL-CI-CT B Bodt (5 copies)
ATTN AMSRL-CI-CT C Hansen
ATTN AMSRL-CI-CT E Heilman
ATTN AMSRL-CI-CT G Moss
ATTN AMSRL-CI-CT J Forester
ATTN AMSRL-CI-CT J O'May
ATTN AMSRL-CI-CT M Thomas
ATTN AMSRL-CI-CT P Jones
ATTN AMSRL-CI-CT R C Kaste
ATTN AMSRL-CI-LP (305)
Aberdeen Proving Ground MD 21005

Director
US Army Rsrch Lab
ATTN AMSRL-RO-D JCI Chang
PO Box 12211
Research Triangle Park NC 27709

US Army Rsrch Lab
ATTN AMSRL-CI-CB P David (3 copies)
ATTN AMSRL-CI-CN D Hillis (5 copies)
ATTN AMSRL-CI-IS-R Mail & Records Mgmt
ATTN AMSRL-CI-IS-T Techl Pub
ATTN AMSRL-CI-OK-TL Techl Lib (2 copies)
ATTN AMSRL-D D R Smith

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2001	Final, FY98

4. TITLE AND SUBTITLE An Application of Statistics to Algorithm Development in Image Analysis

5. FUNDING NUMBERS
DA PR: AH48
PE: 611102H48

6. AUTHOR(S) Barry Bodt, Philip David, and David Hillis

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
Attn: AMSRL-CI-CD email: babodt@arl.army.mil
2800 Powder Mill Road
Adelphi, MD 20783-1197

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-2333

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Ballistic Missile Defense Organization
7100 Defense, Pentagon
Washington, DC 20301-7100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
ARL PR: 1FEPC1
AMS code: 611102H4811

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We gathered and processed infrared (IR) images taken by a quantum well infrared photodetector (QWIP) camera to support the development of an algorithm for airborne target detection. This report addresses the statistical aspects of this effort. Digitized images consisted of background clutter alone and background clutter with a small aircraft target. We used a template-matching technique to detect the small aircraft amidst the background clutter. We propose a reasonable statistic for use in our algorithm. Our primary focus is to describe the experimental design and analysis used in determining parameter settings for the detection algorithm and to report those results.

14. SUBJECT TERMS infrared sensor, target detection, experimental design

15. NUMBER OF PAGES 25

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102